WHAT IS CLAIMED IS:

Selection of D-able subsites

1 1. A method of selecting a target site within a target sequence for
2 targeting by a zinc finger protein comprising:
3 providing a target nucleic acid to be targeted by a zinc finger protein;
4 outputting a target site within the target nucleic acid comprising 5'NNx
5 aNy bNzc3', wherein

6 each of (x, a), (y, b) and (z, c) is (N, N) or (G, K);

7 at least one of (x, a), (y, b) and (z, c) is (G, K). and

8 N and K are IUPAC-IUB ambiguity codes.

1 2. The method of claim 1, further comprising selecting a plurality of
2 potential target sites within the target nucleic acid and outputting a subset of the plurality
3 of potential target segments comprising 5'NNx aNy bNzc3', wherein

4 each of (x, a), (y, b) and (z, c) is (N, N) or (G, K);

5 at least one of (x, a), (y, b) and (z, c) is (G, K). and

6 N and K are IUPAC-IUB ambiguity codes.

1 3. The method of claim 2, wherein the target nucleic acid comprises a
2 target gene.

1 4. The method of claim 1, wherein at least two of (x, a), (y, b) and (z,
2 c) is (G, K).

1 5. The method of claim 1, wherein all three of (x, a), (y, b) and (z, c)
2 are (G, K).

1 6. The method of claim 1, wherein the zinc finger protein comprises
2 three fingers.

1   7.   The method of claim 1, wherein the target site comprises first and

2   second target segments, each comprising 5'NNx aNy bNzc3', and the method further

3   comprises selecting the second target segment.

1   8.   The method of claim 7, wherein in the second segment at least two

2   of (x, a), (y, b) and (z, c) are (G, K).

1   9.   The method of claim 8, wherein in the second segment all three of

2   (x, a), (y, b) and (z, c) are (G, K).

1   10.   The method of claim 10, wherein the first and second target

2   segment are separated by fewer than 5 bases in the target site.

1   11. The method of claim 10, wherein the first target segment comprises

2   5'NNN NNN NNG3', the second target segment comprises 5'KNx aNY bNzc3' and

3   there are zero bases separating the first and second target segments in the target site.

1   12.   The method of claim 7, further comprising synthesizing step

2   comprises synthesizing a first zinc finger protein comprising three zinc fingers that

3   respectively bind to the NNx aNy and bNz triplets in the target segment and a second

4   three fingers that respectively bind to the NNx aNy and bNz triplets in the second target

5   segment.

1   13.   The method of claim 1, further comprising synthesizing a zinc

2   finger protein comprising first, second and third fingers that bind to the bNz aNy and

3   NNx triplets respectively.

1   14. The method of claim 13, wherein each of the first, second and third

2   fingers is selected or designed independently.

1   15. The method of claim 13, wherein a finger is designed from a database

2   containing designation of zinc finger proteins, subdesignations of finger components, and

3   nucleic acid sequences bound by the zinc finger proteins.

16.   The method of claim 13, wherein a finger is selected by screening variants of a zinc finger binding protein for specific binding to the target site to identify a variant that binds to the target site.

17.   The method of claim 13, further comprising contacting a sample containing the target nucleic acid with the zinc finger protein, whereby the zinc finger protein binds to the target site revealing the presence of the target nucleic acid or a particular allelic form thereof.

18. The method of claim 13, further comprising contacting a sample containing the target nucleic acid with the zinc finger protein, whereby the zinc finger protein binds to the target site thereby modulating expression of the target nucleic acid

19.   The method of claim 1, wherein the target site occurs in a coding region of a gene

20.   The method of claim 1, wherein the target site occurs within or proximal to a promoter, enhancer, or transcription start sit

21. The method of claim 1, wherein the target site occurs outside a promoter, regulatory sequence or transcriptional start site within the target nucleic acid.

Selection of Target Sites Using a Correspondence Regim

22.   A method for selecting a target site within a polynucleotide for targeting by a zinc finger protein, comprising:

providing a polynucleotide sequence;

selecting a potential target site of within the polynucleotide sequence; the potential target site comprising contiguous first, second and third triplets of bases at first, second and third positions in the potential target site;

determining a plurality of subscores by applying a correspondence regime between triplets and triplet position in a sequence of three contiguous triplets, wherein each triplet has first, second and third corresponding positions, and each combination of triplet and triplet position has a particular subscore

11   calculating a score for the potential target site by combining subscores for

12   the first, second, and third triplets;

13   repeating the selecting, determining and calculating steps at least once on a

14   further potential target site comprising first, second and third triplets at first, second and

15   third positions of the further potential target site to determine a further score;

16   providing output of at least one potential target site with its score.

1   23.   The method of claim 22, wherein output is provided of the

2   potential target site with the highest score.

1   24.   The method of claim 22, wherein output is provided of the n

2   potential target sites with the highest scores, and the method further comprises providing

3   user input of a value for n.

1   25.   The method of claim 22, wherein the subscores are combined by

2   forming the product of the subscores.

1   26.   The method of claim 25, wherein the correspondence regime

2   comprises 64 triplets, each having first, second, and third corresponding positions, and

3   192 subscores.

1   27.   The method of claim 22, wherein the subscores in the

2   correspondence regime are determined by assigning a first value as the subscore of a

3   subset of triplets and corresponding positions, for each of which there is an existing zinc

4   finger protein that comprising a finger that specifically binds to the triplet from the same

5   position in the existing zinc finger protein as the corresponding position of the triplet in

6   the correspondence regime; assigning a second value as the subscore of a subset of

7   triplets and corresponding positions, for each of which there is an existing zinc finger

8   protein that comprises a finger that specifically binds to the triplet from a different

9   position in the existing zinc finger protein than the corresponding position of the triplet in

10   the correspondence regime; and assigning a third value as the subscore of a subset of

11   triplets and corresponding positions for which there is no known zinc protein comprising

12   a finger that specifically binds to the triplet.

1   28.   The method of claim 22, wherein the correspondence regime is
2   shown in Table 1.

1   29.   The method of claim 22, further comprising combining a context
2   parameter with the subscore of at least one of the first, second and third triplets to give a
3   scaled subscore of the at least one triplet.

1   30.   The method of claim 29, wherein the context parameter is
2   combined with the subscore when the target site comprises a base sequence 5'NNGK3',
3   wherein NNG is the at least one triplet.

1   31.   The method of claim 22, further comprising combining a context
2   parameter that is combined with the score of a potential target site to give a scaled score

1   32.   The method of claim 31, wherein the context parameter is
2   combined with the score when a potential target site comprises 5'NNx aNy bNzc3',
3   wherein

4   wherein each of (x, a), (y, b) and (z, c) is (N, N) or (G, K);

5   at least one of (x, a), (y, b) and (z, c) is (G, K). and

6   N and K are IUPAC-IUB ambiguity codes.

1   33.   The method of claim 32, wherein a first context parameter is
2   combined with the score if one of (x, a), (y, b) and (z, c) is (G, K), and a second context
3   parameter is combined with the score if two of (x, a), (y, b) and (z, c) are (G, K), and a
4   third context parameter is input if three of (x, a), (y, b) and (z, c) are (G, K)

1   34.   The method of claim 22, wherein output is provided of at least a
2   nonoverlapping pair of potential target sites and their scores, the members of the pair
3   being separated by five or fewer bases in the polynucleotide.

Design of ZFPs using a Database

1   35.   A method of producing a zinc finger protein comprising:

2   (a) providing a database comprising designations for a plurality of zinc

3   finger proteins, each protein comprising at least first, second and third fingers, and

4   subdesignations for each of the three fingers of each of the zinc finger proteins;

5   a corresponding nucleic acid sequence for each zinc finger protein, each

6   sequence comprising at least first, second and third triplets specifically bound by the at

7   least first, second and third fingers respectively in each zinc finger protein, the first,

8   second and third triplets being arranged in the nucleic acid sequence (3'-5') in the same

9   respective order as the first, second and third fingers are arranged in the zinc finger

10   protein (N-terminal to C-terminal);

11   (b) providing a target site for design of a zinc finger protein, the target site

12   comprising continuous first, second and third triplets in a 3'-5' order,

13   (c) for the first, second and third triplet in the target site, identifying first,

14   second and third sets of zinc finger protein(s) in the database, the first set comprising zinc

15   finger protein(s) comprising a finger specifically binding to the first triplet in the target

16   site, the second set comprising zinc finger protein(s) comprising a finger specifically

17   binding to the second triplet in the target site, the third set comprising zinc finger

18   protein(s) comprising a finger specifically binding to the third triplet in the target site;

19   (d) outputting designations and subdesignations of the zinc finger proteins

20   in the first, second, and third sets identified in step (c).

1   36.   The method of claim 35, further comprising:

2   (e) producing a zinc finger protein that binds to the target site comprising

3   a first finger from a zinc finger protein from the first set, a second finger from a zinc

4   finger protein from the second set, and a third finger from a zinc finger protein from the

5   third set.

        37.   The method of claim 36 further comprising identifying subsets of

2   the first, second and third sets, the subset of the first set comprising zinc finger protein(s)

3   comprising a finger that specifically binds to the first triplet in the target site from the

4   first finger position of a zinc finger protein in the database; the subset of the second set

5   comprising zinc finger protein(s) comprising a finger that specifically binds to the second

6   triplet in the target site from the second finger position in a zinc finger protein in the

7   database; the subset of the third set comprising zinc finger protein(s) comprising a finger

8   that specifically binds to the third triplet in the target site from a third finger position in a

9   zinc finger protein in the database;

10   wherein

11   the outputting step comprising outputting designations and

12   subdesignations of the subset of the first, second and third sets; and

13   the producing step comprising producing a zinc finger protein comprising

14   a first finger from the first subset, a second finger from the second subset, and a third

15   finger from the third subset.

1   38.   The method of claim 37, wherein the outputting comprises

2   outputting the designations and subdesignations of the subsets of the first, second and

3   third sets, and the first, second and third sets minus their respective subsets.

1   39.   The method of claim 38, wherein each of the subsets is a null set.

1   40.   The method of claim 35, wherein the target site is provided by user

2   input.

1   41.   The method of claim 35 wherein the target site is provided by the

2   method of claim 1 or claim 22.

1   42.   A method of producing a zinc finger protein comprising:

2   (a) providing a database comprising

3   designations for a plurality of zinc finger proteins, each

4   protein comprising at least first, and second fingers,

5   subdesignations for each of the fingers of each of the zinc

6   finger proteins; and

7   a corresponding nucleic acid sequence for each zinc finger protein, each

8   sequence comprising first and second triplets specifically bound by the first and second

9   fingers respectively, the triplets being arranged in the nucleic acid sequence (3'-5') in the

10   same respective order as the first and second and fingers are arranged in the zinc finger

11   protein (N-terminal to C-terminal);

12   (b) providing a target site for design of a zinc finger protein, the

13   target site comprising contiguous first, and second triplets ordered 3'5' in the target site;

14   (c) for the first and second triplet in the target site, identifying first

15   and second sets of zinc finger protein(s) in the database, the first set comprising zinc

16   finger protein(s) comprising a finger specifically binding to the first triplet in the target

17   site, the second set comprising zinc finger protein comprising a finger specifically

18   binding to the second triplet in the target site;

19   (d) outputting designations and subdesignations of the zinc finger proteins in the

20   first, and second sets identified in step (c).

1   43.   A method of producing a zinc finger protein comprising:

2   (a) providing a database comprising:

3   designations for a plurality of zinc finger proteins, each

4   protein comprising at least first, and second fingers;

5   subdesignations for each of the fingers of each of the zinc

6   finger proteins; and

7   a corresponding nucleic acid sequence for each zinc finger

8   protein, each sequence comprising first, and second triplets specifically bound by the first

9   and second fingers respectively, the triplets being arranged in the nucleic acid sequence

10   (3'-5') in the same respective order as the first and second and fingers are arranged in the

11   zinc finger protein (N-terminal to C-terminal);

12   (b) providing a target site for design of a zinc finger protein, the

13   target site comprising contiguous first, second and third triplets ordered 3'5' in the target

14   site;

15   (c) for the first and third triplet in the target site, identifying first

16   and second sets of zinc finger protein(s) in the database, the first set comprising zinc

17   finger protein(s) comprising a finger specifically binding to the first triplet in the target

18   site, the second set comprising zinc finger protein comprising a finger specifically

19   binding to the third triplet in the target site;

20   (d) outputting designations and subdesignations of the zinc finger proteins in the

21   first, and second sets identified in step (c).

1   44.   A computer program product for selecting a target sequence within

2   a polynucleotide for targeting by a zinc finger protein, comprising:

3   (a) code for providing a polynucleotide sequence;

4   (b) code for selecting a potential target site within the polynucleotide

5   sequence; the potential target site comprising first, second and third triplets of bases at

6   first, second and third positions in the potential target site;

7   (c) code for calculating a score for the potential target site from a

8   combination of subscores for the first, second, and third triplets, the subscores being

9   obtained from a correspondence regime between triplets and triplet position, wherein each

10   triplet has first, second and third corresponding positions, and each corresponding triplet

11   and position has a particular subscore;

12   (d) code for repeating steps (b) and (c) at least once on a further potential

13   target site comprising first, second and third triplets at first, second and third positions of

14   the further potential target site to determine a further score;

15   (e) code for providing output of at least one of the potential target site

16   with its score

17   (f) a computer readable storage medium for holding the codes

1   45. The computer program product of claim 44, further comprising code

2   for combining a context parameter with a subscore.

1   46. A system for selecting a target sequence within a polynucleotide

2   for targeting by a zinc finger protein, comprising:

3   (a) a memory;

4   (b) a system bus;

5   (c) a processor operatively disposed to:

6   (1) provide or receive a polynucleotide sequence;

7   (2) select a potential target site within the polynucleotide sequence; the

8   potential target site comprising first, second and third triplets of bases at first, second and

9   third positions in the potential target site;

10   (3) calculate a score for the potential target site from a combination of

11   subscores for the first, second, and third triplets, the subscores being obtained from a

12   correspondence regime between triplets and triplet position, wherein each triplet has first,

13   second and third corresponding positions, and each corresponding triplet and position has

14   a particular subscore;

15   (4) repeat steps (2) and (3) at least once on a further potential target site

16 comprising first, second and third triplets at first, second and third positions of the further

17 potential target site to determine a further score;

18   (5) provide output of at least one of the potential target site with its score.

1   47. The system of claim 46, wherein the processor is further operatively

2 disposed to combine a context parameter with a subscore.

1   48.   A computer program product for designing a zinc finger protein

2 comprising:

3   (a) code for providing a database comprising

4   designations for a plurality of zinc finger proteins, each protein

5 comprising at least first, second and third fingers ,

6   subdesignations for each of the three fingers of each of the zinc

7 finger proteins;

8   a corresponding nucleic acid sequence for each zinc finger protein,

9 each sequence comprising at least first, second and third triplets specifically bound by the

10 at least first, second and third fingers respectively in each zinc finger protein, the first,

11 second and third triplets being arranged in the nucleic acid sequence (3'-5') in the same

12 respective order as the first, second and third fingers are arranged in the zinc finger

13 protein (N-terminus to C-terminus);

14   (b) code for providing a target site for design of a zinc finger protein, the

15 target site comprising at least first, second and third triplets,

16   (c) for the first, second and third triplet in the target site, code for

17 identifying first, second and third sets of zinc finger protein(s) in the database, the first set

18 comprising zinc finger protein(s) comprising a finger specifically binding to the first

19 triplet in the target site, the second set comprising a finger specifically binding to the

20 second triplet in the target site, the third set comprising a finger specifically binding to the

21 third triplet in the target site;

22   (d) code for outputting designations and subdesignations of the zinc finger

23 proteins in the first, second, and third sets identified in step (c).

24   (e) a compute readable storage medium for holding the codes.

1   49.   A system for designing a zinc finger protein comprising:

2   (a) a memory;

3   (b) a system bus;

4   (c) a processor operatively disposed to:

5   (1) provide a database comprising

6   designations for a plurality of zinc finger proteins, each protein comprising

7   at least first, second and third fingers, subdesignations for each of the three fingers of

8   each of the zinc finger proteins;

9   a corresponding nucleic acid sequence for each zinc finger protein, each

10   sequence comprising at least first, second and third triplets specifically bound by the at

11   least first, second and third fingers respectively in each zinc finger protein, the first,

12   second and third triplets being arranged in the nucleic acid sequence (3'-5')in the same

13   respective order as the first, second and third fingers are arranged in the zinc finger

14   protein (N-terminus to C-terminus);

15   (2) provide or be provided with a target site for design of a zinc

16   finger protein, the target site comprising at least first, second and third triplets,

17   (3) for the first, second and third triplet in the target site,

18   identifying first, second and third sets of zinc finger protein(s) in the database, the first set

19   comprising zinc finger protein(s) comprising a finger specifically binding to the first

20   triplet in the target site, the second set comprising a finger specifically binding to the

21   second triplet in the target site, the third set comprising a finger specifically binding to the

22   third triplet in the target site;

23   (4) output designations and subdesignations of the zinc finger

24   proteins in the first, second, and third sets identified in step (3).

1   50.   A computer program product for selecting a target site within a

2   target sequence for targeting by a zinc finger protein comprising:

3   code for providing a target nucleic acid to be targeted by a zinc finger

4   protein;

5   code for outputting a target site within the target nucleic acid comprising

6   5'NNx aNy bNzc3', wherein α

7   wherein each of (x, a), (y, b) and (z, c) is (N, N) or (G, K);

8   at least one of (x, a), (y, b) and (z, c) is (G, K). and

9   N and K are IUPAC-IUB ambiguity codes;

10   and a computer readable storage medium for holding the codes.

1   51.   A system for selecting a target site within a target sequence for

2   targeting by a zinc finger protein comprising:

3   (a) a memory;

4   (b) a system bus;

5   (c) a processor operatively disposed to:

6   provide a target nucleic acid to be targeted by a zinc finger protein;

7   output a target site within the target nucleic acid comprising 5'NNx aNy

8   bNzc3', wherein

9   wherein each of (x, a), (y, b) and (z, c) is (N, N) or (G, K);

10   at least one of (x, a), (y, b) and (z, c) is (G, K) and

11   N and K are IUPAC-IUB ambiguity codes.

add a58